COGS 118A: ASSIGNMENT 1

Problem 1. Matlab: defining a function and using subplots.

Define a function

\[ h(x; w) = \frac{1}{1 + e^{-wx}} \]

For each of the four values of \( w = \{0, .015, .1, 1000\} \) plot \( h(x; w) \) over \( x = [-70 : 70] \). Using the \texttt{subplot} command, place these plots in a 2-by-2 grid. Title each subplot with the value of \( w \) used in that subplot. Standardize the axes across subplots with the command \texttt{axis} for ease of comparison.

Print out this figure.

Notice from the plots that a single function can behave like a constant, a line, a sigmoid, or a step function, depending on the value of the parameter \( w \). This family of functions is known as logistic functions. You will see them again frequently later in the course (not the least in logistic regression!).

Problem 2. Introduction to regression: polynomial curve-fitting

(a) You will generate data from a quadratic function of the form

\[ y = ax^2 + bx + c \]

First generate 10 values for \( x \) between -10 and 10 using the command \texttt{x = rand(10,1)*20-10}. Then generate the corresponding \( y \)-values when \( a = 1, b = -4, c = -21 \). Now produce a “noisy” version of the \( y \) values, using the \texttt{normrnd} command to generate Gaussian noise with \( \mu = 0, \sigma = 10 \), which you add to \( y \), yielding the noisy \( y_n \).

Use the command \texttt{plot} to plot \( y \) versus \( x \) with a blue line. Use the command \texttt{hold} and \texttt{plot} again to plot \( y_n \) versus \( x \) with green circles.

Use the command \texttt{regress} to find the best 0th-, 1st-, 2nd-, 5th-, and 10th-order polynomial fits to \( y_n \) as a function of \( x \). Use \texttt{plot} to display the polynomial fits with different-colored dashed lines, alongside \( y \) and \( y_n \). Use \texttt{legend} to label \( y \), \( y_n \), and the various polynomial fits.

Print out this figure.

(b) Compute the root-mean-square error for each polynomial fit. In a new figure, plot this error (known as “training error”) against the order of the polynomial. Then generate 10 more values for \( x \), using the command \texttt{x2 = rand(10,1)*20-10}. Compute the “true” \( y_2 \) values using the equation in (a). Now compute the root-mean-square error between the polynomial fit obtained from (a) and \( y_2 \). Use \texttt{hold} and \texttt{plot} to plot this “test error” as a function of the order of the polynomial. How does it compare to the training error’s dependence on the order of the polynomial? Print out this figure.
Problem 3. Math review. Prove that the following are true:

(a) Cyclic permutations are allowed inside the trace:
\[ \text{Tr}(BCD) = \text{Tr}(DBC) = \text{Tr}(CDB) \]

(b) Trace of a scalar is a scalar:
\[ x^\top Bx = \text{Tr}(x^\top Bx) = \text{Tr}(xx^\top B) \]

(c) Some calculus review
\[ \frac{d}{dz} \left[ \frac{1}{(b+cz)} \exp(az) + e \right] = \exp(az) \left( \frac{a}{b+cz} - \frac{c}{(b+cz)^2} \right) \]
Note that \(a, b, c\) and \(e\) are constants and \(1/u = u^{-1}\). Hint: recall the chain rule.

Problem 4. Basic probability theory

(a) Show that closure under complementation and closure under countable union together imply closure under countable intersection:
\[ E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_i E_i \in \mathcal{F} \]

Hint: use the set theory version of De Morgan’s Law (http://en.wikipedia.org/wiki/De_Morgan’s_laws)

(b) Suppose we are casting a die, and we know we can assign a probability to each atomic event (e.g. \(P(x=i) = 1/6\ \forall i \in \{1, 2, 3, 4, 6\}\)). From that we can compute the probability of many other events, like the event of the die coming up odd, or being greater than 5. Including the null event = \(\{}\) and the all-inclusive event of the sample space itself \(\Omega = \{1, 2, 3, 4, 5, 6\}\), how many total measurable events are there?

(c) Monotonicity: for any two events \(A_1, A_2 \in \mathcal{F}\), show that \(A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)\).

(d) Sub-additivity: For any events \(A, A_i \in \mathcal{F}\), show that \(A \subseteq \bigcup_i A_i \Rightarrow P(A) \leq \sum_i P(A_i)\).

(e) Bayes’ Theorem:
\[ P(B|A) = \frac{P(A|B)P(B)}{\sum_b P(A|B=b)P(B=b)} \]

Problem 5. Independence and expectation (assume \(x\) and \(y\) are continuous-valued r.v.’s with pdf’s)

(a) Show that if the r.v.’s \(x\) and \(y\) are independent, then \(p(x|y) = p(x)\) and \(p(y|x) = p(y)\).

(b) Demonstrate linearity of expectation: \[ \mathbb{E}[af_1(x) + bf_2(x)] = a\mathbb{E}[f_1(x)] + b\mathbb{E}[f_2(x)] \]

(c) Variance is not a linear operation, but still has some nice properties. Show the following is true for two independent r.v. \(x\) and \(y\):
\[ \text{var}[ax + by] = a^2\text{var}[x] + b^2\text{var}[y] \]

(d) Use the definition of expectation for a continuous random variable, compute \(\mathbb{E}[x]\) when \(x\) has a uniform distribution over the unit interval:
\[ p(x) = \begin{cases} 1, & x \in [0, 1] \\ 0, & \text{otherwise} \end{cases} \]

(e) Show that if the random variables \(x\) and \(y\) are independent, then \(\text{cov}[x, y] = 0\).